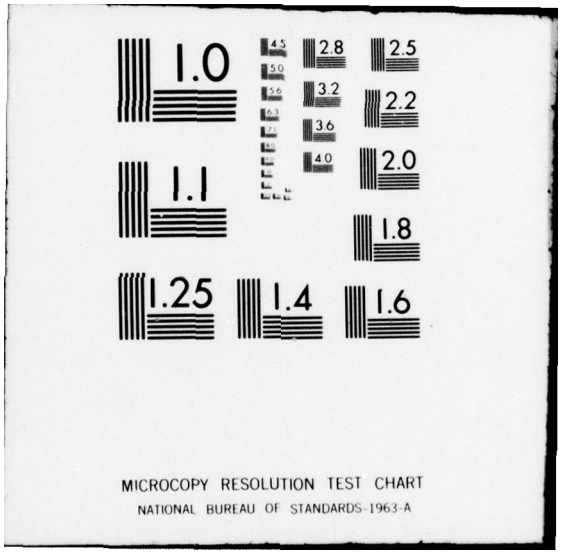
AD-A064 183 NAVAL RESEARCH LAB WASHINGTON D C F/G 9/4
MINIMUM CROSS-ENTROPY SPECTRAL ANALYSIS.(U)
JAN 79 J E SHORE
UNCLASSIFIED NRL-MR-3921 SBIE-AD-E000 261 NL

END
DATE FILMED
4 -79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NRL Memorandum Report 3921

Minimum Cross-Entropy Spectral Analysis

JOHN E. SHORE

Information Systems Staff
Communications Sciences Division

January 16, 1979

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NRL Memorandum Report 3921 (19 NRL-MR-)
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): (6) MINIMUM CROSS-ENTROPY SPECTRAL ANALYSIS,
5. TYPE OF REPORT & PERIOD COVERED: (9) Interim report on a continuing NRL problem
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): (10) John E. Shore
8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Research Laboratory, Washington, D.C. 20375
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NRL Problem B02-35, 61153N, RR014-09-41

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, D.C. 22217 (11 16)
12. REPORT DATE: January 16, 1979
13. NUMBER OF PAGES: 35

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12 36p.) (16 RR0140 9)
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited. (17 RR0140941) (18 SBIE)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): (19 AD-E000 261)

DDC RECEIVED FEB 5 1979 B

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spectral analysis
Maximum entropy
Maximum entropy spectral analysis
Minimum cross-entropy
Directed divergence
Information theory
Discrimination information

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The principle of minimum cross-entropy (minimum directed divergence) is summarized, discussed, and applied to the classical problem of estimating power spectra given samples of the autocorrelation function. This new approach reduces to maximum entropy spectral analysis (MESA) in certain special cases, and thereby provides a fundamental derivation of MESA. In contrast to MESA, the minimum cross-entropy approach makes use of prior information about the power spectrum. Depending on the extent of prior information, various alternative minimum

(Continues)

DD FORM 1473 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 950

20. Abstract (Continued)

cross-entropy spectral estimates are obtained. When a prior estimate of the power spectrum is available, the minimum cross-entropy result differs from the MESA result. Results are derived in two equivalent ways; once by minimizing the cross-entropy of underlying probability densities, and once by arguments concerning the cross-entropy between the input and output of linear filters.

CONTENTS

I. INTRODUCTION 1

A A. Maximum Entropy Spectral Analysis (MESA) 1
B. Limited Acceptance of MESA Viewpoint 3
C. Outline 4

II. CROSS ENTROPY MINIMIZATION 5

A. A General Inference Problem Involving Probability Densities 5
B. The Principle of Minimum Cross-Entropy 6
C. Background and Justification of Cross-Entropy Minimization 8

III. MINIMUM CROSS-ENTROPY PROBABILITY DENSITIES FOR STOCHASTIC SIGNALS 10

A. Power Spectrum Probability Densities 10
B. Minimum Cross-Entropy Densities Given Expected Spectral Powers 10
C. Spectral Probability Densities for White Noise 14

IV. CROSS-ENTROPY BETWEEN INPUT AND OUTPUT OF LINEAR FILTER 16

V. MINIMUM CROSS-ENTROPY POWER SPECTRA GIVEN AUTO-CORRELATION INFORMATION 17

A. Results When No Prior Power Spectrum Estimate is Given 18
B. Results When A Prior Power Spectrum Estimate is Given 21

VI. COMPARISON OF RESULTS 23

VII. TOWARDS EFFICIENT ALGORITHMS 26

VIII. CONCLUSIONS 28

IX. ACKNOWLEDGMENT 29

REFERENCES 30

I. INTRODUCTION

Work reported in /1/ has shown that the principle of minimum cross-entropy (minimum directed divergence) provides a correct, general method of inductive inference in terms of continuous probability densities when given a prior density and information about the "true" density in the form of expected values. In this paper, I show how cross-entropy minimization can be used to estimate power spectra when given a prior estimate of the spectrum and new information in the form of autocorrelation function samples. This new approach reduces to maximum entropy spectral analysis /2/ in certain special cases, and can be thought of as providing a fundamental derivation of the maximum entropy technique.

A. Maximum Entropy Spectral Analysis (MESA)

Because the power spectrum S(f) of a band-limited, stationary process is related to its autocorrelation function R(t) by a Fourier transform, and because it is relatively easy to measure the autocorrelation function, many spectral analysis techniques start with samples of the autocorrelation function. The classical approach uses spectral window functions /3/. In this approach one takes the Fourier transform of the product R(t)W(t), where R(t) is the measured autocorrelation function in the range $|t| < T$, and where W(t) is a known window function with $W(t) = 0$ for $|t| > T$. One then estimates the unknown power spectrum S(f) by exploiting the convolution theorem, which states that the Fourier transform of the product of two time domain functions is equal to the convolution in the frequency domain of their Fourier transforms. Although mathematically elegant, the classical procedure can be

Note: Manuscript submitted November 29, 1978.

seen to distort the known values of R(t), $|t| < T$, and to assume that R(t) = 0 in the unknown region $|t| > T$, despite the fact that R(t) cannot in general be zero everywhere in this region. An alternative approach is to extend R(t) so as to take on reasonable values in the unknown region $|t| > T$ and to estimate S(f) by taking the Fourier transform of the resulting extended function. As a general approach this seems more reasonable than the classical approach, but it leaves open the question of how to extend the measured portion of R(t).

In proposing the technique called Maximum Entropy Spectral Analysis (MESA), Burg /2/ suggested that R(t) be extended in a manner that maximizes the entropy of the underlying stationary process. Specifically, Burg proposed that the power spectrum S(f) be estimated by maximizing

$$\int_0^W df \, \log(S(f)) \tag{1}$$

subject to the known constraints

$$R(t_k) = \int_{-W}^{W} df \, S(f) \exp(2\pi i t_k f), \tag{2}$$

where W is the bandwidth, and where $R(t_k)$, k=1,2,...,m, are known samples of the autocorrelation function.

Maximum entropy spectral analysis can be seen /4/ as an application of Jaynes's maximum entropy principle /5/, which applies to situations in which one wishes to estimate or guess at an unknown probabilities $q^t(x_i)$ when given a set of expected values $\bar{g}_k$

$$\bar{g}_k = \sum_i q^t(x_i) g_k(x_i) \, , \tag{3}$$

k = 1,2,...,m. The maximum entropy principle states that, of all the distributions that satisfy the constraints (3), one should choose the one with

the largest entropy

$$H(q) = - \sum_i q(x_i)\log(q(x_i)) \quad . \tag{4}$$

Intuitively, the maximum entropy principle follows from the the fact that, to within the choice of logarithmic base, entropy (4) is a unique measure of the uncertainty represented by the distribution $q(x_i)$ /6/,/7/. Jaynes argues that the maximum entropy distribution is "the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have" /5, p. 623/. Similarly, the maximum entropy distribution "agrees with what is known, but expresses 'maximum uncertainty' with respect to all other matters" /8, p.231/.

The maximum entropy principle is applied somewhat indirectly in MESA. The expression (1) is the entropy gain in a stochastic process that is passed through a linear filter with characteristic function Y(f), where $S(f) = |Y(f)|^2$ (see /6, pp. 93-95/, /14, pp.412-414/, /26, p. 243/). If the input process is white noise, then the output process has spectral power density S(f). This suggests that the process entropy can be maximized by maximizing the entropy gain of the filter that produces the process. Thus, (1) is maximized subject to the constraints (2).

B. Limited Acceptance of MESA Viewpoint

Burg's proposal /2/ led to a variety of practical and useful spectral estimation algorithms /9/-/18/, but it seems fair to say that, despite its strong intuitive appeal, MESA has not had widespread acceptance. The reasons for this appear to go beyond the natural inertia that results from familiarity with the long-standing, classical approach, particularly since MESA is known to be equivalent to minimum least-squares estimation /10/,/19/.

I believe that much of the resistance to the MESA viewpoint stems from doubt about the validity of the maximum entropy principle, which has remained controversial /20/-/25/ despite numerous successful applications (see /1/). To some, entropy's properties as an information measure make it obvious that entropy maximization is the correct way to account for constraint information. To others, such an informal and intuitive justification yields plausibility for the maximum entropy principle, but not proof --- why maximize entropy, why not some other function? Moreover, even if one accepts the maximum entropy principle, there are well-known problems /7/ with extending it from (3)-(4) to the continuous case. Such an extension is required, since derivations of (1) deal with continuous probability densities (/6, pp. 93-95/, /14, pp. 412-14/, /26, p. 243/). Some of the resistance to MESA may also stem from the fact that the maximum entropy principle is applied indirectly in terms of filtering rather than directly in terms of underlying probability densities.

All of these hesitations can be addressed in light of the results for cross-entropy minimization that were obtained in /1/.

C. Outline

Section II summarizes the principle of minimum cross-entropy and discusses informally the sense in which this principle provides a correct, general method of inductive inference /1/. In Section III, I describe stochastic signals in terms of frequency domain probability densities, I derive the minimum cross-entropy density given known expected spectral powers, and I discuss two different possible densities for white noise -- one uniform probability density and one non-uniform density. In Section IV, I derive the cross-entropy between the input and output of a linear filter and show that

the resulting expression reduces to (1) when the input is one of the white noise densities introduced in Section III but not when the input is the other one. Section V contains derivations of minimum cross-entropy densities and corresponding power spectrum estimates when given information in the form of autocorrelation samples for cases both with and without previous estimates of the power spectrum. The derivations are carried out twice --- once directly in terms of the underlying probability densities, and once indirectly in terms of linear filters. The results are compared with those of MESA in Section VI. Some remarks about possible algorithms (Section VII) are then followed by brief conclusions (Section VIII).

II. CROSS ENTROPY MINIMIZATION

A. A General Inference Problem Involving Probability Densities

Let $\underset{\sim}{x}$ denote a single state of some system that has a set $\underset{\sim}{D}$ of possible system states and a probability density $q^{\dagger}(\underset{\sim}{x})$ of states. Let $\mathbb{Q}$ be the set of all probability densities q on $\underset{\sim}{D}$ such that $q(\underset{\sim}{x}) \gg 0$ for $\underset{\sim}{x} \in \underset{\sim}{D}$ and

$$\int_{\underset{\sim}{D}} d\underset{\sim}{x}\, q(\underset{\sim}{x}) = 1 \quad . \tag{5}$$

We assume that the existence of $q^{\dagger} \in \mathbb{Q}$ is known but that $q^{\dagger}$ itself is unknown. The density $q^{\dagger}$ is sometimes known as a "true" density.

Suppose $p \in \mathbb{Q}$ is a prior density that is our current estimate of $q^{\dagger}$, and suppose we gain new information about $q^{\dagger}$ in the form of a set of expected values

$$\int_{\underset{\sim}{D}} d\underset{\sim}{x}\, q^{\dagger}(\underset{\sim}{x}) g_r(\underset{\sim}{x}) \equiv \langle g_r \rangle = \bar{g}_r \ , \tag{6}$$

for a known set of bounded functions $g_r(\underset{\sim}{x})$ and numbers $\bar{g}_r$, r = 1,...,m. Now, because the constraints (6) do not determine $q^\dagger$ completely, they are satisfied not only by $q^\dagger$ but by some subset of densities $\mathcal{S} \subseteq \mathcal{Q}$. Which single density should we choose from this subset to be our new estimate of $q^\dagger$, and how should we use the prior p and the new information (6) in making this choice?

B. The Principle of Minimum Cross-Entropy

The solution to this inference problem is obtained by minimizing a functional H(q,p) called cross-entropy,

$$H(q,p) = \int_{\underset{\sim}{D}} d\underset{\sim}{x}\; q(\underset{\sim}{x}) \log(q(\underset{\sim}{x})/p(\underset{\sim}{x})) \; . \tag{7}$$

Specifically, of all the densities $q' \in \mathcal{S}$ that satisfy the constraints (6), we choose the one with the smallest cross-entropy H(q',p) with respect to the prior p. Stated differently, the posterior density q satisfies

$$H(q,p) = \min_{q' \in \mathcal{S}} H(q',p) \; ,$$

where $\mathcal{S} \subseteq \mathcal{Q}$ comprises all of the densities that satisfy the constraints (6).

Mathematically, the solution is obtained using the method of Lagrangian multipliers and standard techniques from the calculus of variations. The minimization condition is

$$\log(q(\underset{\sim}{x})/p(\underset{\sim}{x})) + 1 + \lambda_o + \sum_r \beta_r g_r(x) = 0 \; , \tag{9}$$

where the β_r are Lagrangian multipliers corresponding to the constraints (6), and where λ_o is a Lagrangian multiplier corresponding to the normalization constraint (5). The solution of (9) is

$$q(\underset{\sim}{x}) = p(\underset{\sim}{x})\exp(-\lambda - \sum_r \beta_r g_r(x)), \tag{10}$$

where $\lambda = \lambda_o + 1$. It is convenient to write (10) in the form

$$q(x) = Z^{-1} p(\underset{\sim}{x})\exp(-\sum_r \beta_r g_r(x)) , \tag{11}$$

where Z is the "partition function",

$$Z = \exp(\lambda) = \int_{\underset{\sim}{D}} d\underset{\sim}{x}\; p(\underset{\sim}{x})\exp(-\sum_r \beta_r g_r(\underset{\sim}{x})) . \tag{12}$$

The values of the multipliers β_r are determined by the known expectation values $\bar{g}_r$ in (6). One can express the posterior q directly in terms of the values $\bar{g}_r$ by solving the equations

$$g_r = Z^{-1}\frac{\partial Z}{\partial \beta_r}$$

$$= -\frac{\partial}{\partial \beta_r}\log(Z) \tag{13}$$

for the β_r, or by substituting (11) into the constraint equations (6) and solving for the β_r. Such solutions are often difficult or impossible to obtain analytically, but one can obtain them computationally in general /1, Appendix B/, /27/.

The principle of minimum cross entropy also applies when, in addition to equality constraints (6), we gain new information about $q^\dagger$ in the form of a bound on an expected value,

$$\int_{\underset{\sim}{D}} d\underset{\sim}{x}\; q^\dagger(\underset{\sim}{x})g(\underset{\sim}{x}) \equiv \langle g \rangle \gg \bar{g} . \tag{14}$$

Such an inequality constraint is handled as follows: First one solves for the minimum cross-entropy density given only the equality constraints (6). If the

resulting density happens to satisfy (14), then this density is the overall solution. If (14) is not satisfied, then the overall solution is the minimum cross-entropy density given (6) and the additional equality constraint $\langle g \rangle = \overline{g}$.

C. Background and Justification of Cross-Entropy Minimization

Cross-entropy goes by other names, including expected weight of evidence /28, p. 72/, directed divergence /29, p. 6/, and relative entropy /20/. The term cross-entropy is due to Good /30/. The principle of minimum cross-entropy was first proposed by Kullback /29, p.37/, who called it a principle of minimum directed divergence or minimum discrimination information. It has been advocated in various forms by others /30/, /31/, /32/, including Jaynes /8/, /33/, who showed that generalizing entropy maximization to continuous densities leads to (7) with p(x) being called an "invariant measure" instead of a prior density. Since entropy maximization does not deal with prior densities --- there being an implicit assumption of uniform priors --- this just expresses the fact that a uniform prior in one coordinate system may not be uniform in another. Cross-entropy minimization has been applied primarily to statistics /29/, /30/, /34/, but also to statistical mechanics /35/, chemistry /36/, pattern recognition /37/, /38/, and the computer storage of probability distributions /39/.

Like entropy, cross-entropy can be characterized axiomatically /32/. Its properties are desirable for an information measure /31/, /32/, and it can be argued /40/ that cross-entropy measures the amount of information necessary to change a prior p into the posterior q. The principle of cross-entropy minimization then follows intuitively, much like entropy maximization. In /1/ we argued that such justifications are weak, not only because they rely on informal, intuitive arguments, but also because they are indirect --- they are

based on a formal description of what is required of an information measure rather than on a formal description of what is required of a method for taking new information into account.

Our approach in /1/ was to formalize the requirements of inductive inference directly in terms of a set of consistency axioms that make no reference to information measures or properties of information measures. All of the axioms are based on a single fundamental principle: If a problem can be solved in more than one way, the results should be consistent. Informally, the axioms may be phrased as follows:

1) Uniqueness. The results of taking new information into account should be unique.

2) Invariance. It shouldn't matter in which coordinate system we account for new information.

3) System independence. It shouldn't matter whether we account for independent information about independent systems separately in terms of different probability densities or together in terms of a joint density.

4) Subset Independence. It shouldn't matter whether we account for information about an independent subset of system states in terms of a separate conditional density or in terms of the full system density.

We were then able to prove /1/ that the principle of minimum cross-entropy provides a correct, general method of inductive inference in the following sense: Given a prior density and new information in the form of constraints on expected values, there is only one posterior density satisfying these constraints that can be chosen in a manner that satisfies the axioms; this unique posterior can be obtained by minimizing cross-entropy.

III. MINIMUM CROSS-ENTROPY PROBABILITY DENSITIES FOR STOCHASTIC SIGNALS

A. Power Spectrum Probability Densities

Consider time-domain signals of the form

$$s(t) = \sum_{k=1}^{n} a_k \cos(\omega_k t) + b_k \sin(\omega_k t) \quad , \tag{15}$$

with non-zero ω_k that need not be uniformly spaced. These are discrete-spectrum, band-limited signals without DC components. (The assumption of no DC term, which is reasonable for many signal processing applications, is made for mathematical convenience.) The power at each frequency is given by the variables x_k,

$$x_k = a_k^2 + b_k^2 \quad . \tag{16}$$

If we consider the x_k to be random variables, we may describe a stochastic signal in terms of a joint probability density $q(\underset{\sim}{x})$, where we write $\underset{\sim}{x}$ for $x_1, x_2, \ldots, x_n$. Instead of constantly referring to $q(\underset{\sim}{x})$ as the spectral power probability density of a stochastic signal, we will informally refer to $q(\underset{\sim}{x})$ as a "signal."

B. Minimum Cross-Entropy Densities Given Expected Spectral Powers

Consider first the problem of choosing $q(\underset{\sim}{x})$ when we know the total expected power per discrete frequency

$$P = \frac{1}{n} \int_{\underset{\sim}{D}} d\underset{\sim}{x} \, (\sum\nolimits_k x_k) q(\underset{\sim}{x}) \quad , \tag{17}$$

where $d\underset{\sim}{x} = dx_1 dx_2 \ldots dx_n$. To apply the principle of minimum cross-entropy, we need a prior density $p(\underset{\sim}{x})$ to represent our state of knowledge before we learn even (17). Since in any real situation there will

be a physical limit on the magitude of the x_k, we assume that the domain of $\underset{\sim}{x}$ is bounded. We may therefore use a uniform prior density. In general, whether or not it is valid to assume a uniform prior density for continuous probablity densities is a difficult question /8/. Therefore, although we assume a uniform prior $p(\underset{\sim}{x})$ in the following, we shall consider the question more carefully later in this section.

We choose $q(\underset{\sim}{x})$ by minimizing cross-entropy subject to the constraints (5) and (16). The result (see (10)) is

$$q(\underset{\sim}{x}) = A \exp(-\beta \sum_k x_k) \quad ,$$

where β is the Lagrangian multiplier corresponding to (16), and where the uniform prior and the Lagrangian multiplier corresponding to (5) have been absorbed into the constant A,

$$A^{-1} = \int dx_1 \int dx_2 \int \cdots \int dx_n \exp(-\beta \sum_k x_k) \quad . \tag{18}$$

Provided that P is much less than the maximum value of the x_k, we may use integration limits of $(0, \infty)$ in (18); this leads to $A = \beta^n$. In terms of the multiplier β, the total power constraint (17) becomes

$$P = \frac{\beta^n}{n} \sum_k \int dx_k \; x_k \exp(-\beta x_k) \prod_{m \neq k} dx_m \exp(-\beta x_m)$$

$$= 1/\beta$$

The posterior $q(\underset{\sim}{x})$ is therefore

$$q(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/P) \exp(-x_k/P) \quad . \tag{19}$$

Thus, $q(\underset{\sim}{x})$ is a multivariate exponential --- each spectral power x_k is

exponentially distributed with mean P.

Now consider the more general problem in which we learn the expected spectral power P_k at each frequency,

$$P_k = \langle x_k \rangle = \int d\underset{\sim}{x}\, x_k q(\underset{\sim}{x}) \quad . \tag{20}$$

Again using a uniform prior, the minimum cross-entropy posterior in this case is

$$q(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/P_k) \exp(-x_k/P_k) \tag{21}$$

(the derivation is similar to that of (19)). In fact, the same posterior (21) results if (19) is used as a prior instead of a uniform prior.

We now return to the question of the uniform prior. One might wonder how differently (21) might have turned out had we described the signal originally as a probability density in the variables a_k, b_k (see (15)) and used a prior that was uniform in a_k, b_k rather than uniform in the variables x_k (see (16)). In this case the constraints (20) take on the form

$$P_k = \langle a_k^2 + b_k^2 \rangle = \int d\underset{\sim}{a} d\underset{\sim}{b}\, (a_k^2 + b_k^2) q(\underset{\sim}{a},\underset{\sim}{b}) \quad . \tag{22}$$

With a uniform prior $p(\underset{\sim}{a},\underset{\sim}{b})$, the result of cross-entropy minimization is

$$q(\underset{\sim}{a},\underset{\sim}{b}) = A \exp(-\textstyle\sum_k \beta_k (a_k^2 + b_k^2)) \quad .$$

Solving for A, using (5), and for the multipliers β_k, using (22), yields

$$q(\underset{\sim}{a},\underset{\sim}{b}) = \prod_{k=1}^{n} (1/\pi P_k) \exp(-(a_k^2 + b_k^2)/P_k) \tag{23}$$

Thus the variables a_k and b_k have Gaussian distributions with zero means and variances $P_k/2$. Since the variances correspond to power expectations

a_k^2 or g_k^2, this just shows that the expected power P_k is divided evenly between the two quadrature components.

Our next step is to transform (23) to a density in terms of of the variables x_k and to compare the result with (21). We begin by transforming from (a_k, b_k) coordinates to (r_k, θ_k) coordinates, where

$$r_k^2 = a_k^2 + b_k^2 \quad ,$$

and

$$\theta_k = \tan^{-1}(b_k/a_k) \quad .$$

The volume elements in the two coordinate systems are related by $da_k db_k = r_k dr_k d\theta_k$. Since $q(\underset{\sim}{a},\underset{\sim}{b})d\underset{\sim}{a}d\underset{\sim}{b} = q(\underset{\sim}{r}, \underset{\sim}{\theta})d\underset{\sim}{r}d\underset{\sim}{\theta}$, it follows that

$$q(\underset{\sim}{r}, \underset{\sim}{\theta}) = \prod_{k=1}^{n} (r_k/\pi P_k)\exp(-r_k^2/P_k)$$

and

$$q(\underset{\sim}{r}) = \prod_{k} (2r_k/P_k)\exp(-r_k^2/P_k) \qquad (24)$$

hold, where in (24) we have integrated over the θ_k coordinates. Now r_k and x_k are related by $x_k = r_k^2$, so that $dx_k = 2r_k dr_k$. Since $q(\underset{\sim}{x})d\underset{\sim}{x} = q(\underset{\sim}{r})d\underset{\sim}{r}$, it follows from (24) that $q(\underset{\sim}{x})$ is given by

$$q(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/P_k)\exp(-x_k/P_k) \quad ,$$

which is the same as (21). Thus, when given information in the form of expected spectral powers, it doesn't matter whether the prior density is assumed to be uniform in the amplitude variables a_k, b_k or in the power variables x_k. The result is a multivariate exponential in the variables x_k or a multivariate Gaussian in the variables a_k, b_k.

One other alternative that needs consideration is to work in the $(\underset{\sim}{r}, \underset{\sim}{\theta})$

coordinates and to use a prior p(r,) that is uniform with respect to the volume element drd . Integrating over and transforming to the x_k coordinates leads to a non-uniform prior

$$p(x) = \frac{1}{2} \prod_{k=1}^{n} (x_k)^{-1/2},$$

This constrasts with the choice of a uniform prior $p(\underset{\sim}{a},\underset{\sim}{b})$, which corresponds to a uniform prior $p(\underset{\sim}{x})$. Since there is no reason to have a non-uniform prior $p(\underset{\sim}{x})$, we reject the possibility of a uniform prior $p(\underset{\sim}{r},\underset{\sim}{\theta})$.

C. Spectral Probability Densities for White Noise

By "white", we mean that the expected spectral powers $\langle x_k \rangle$ are all equal. What probability density should we use to represent white noise? There are two obvious possibilities. The first is the uniform prior $p(\underset{\sim}{x})$ itself, for which $x_k = x_{max}/2$, where x_{max} is the maximum value of the x_k. This is appealing because it doesn't require any additional information beyond the prior. Sometimes, however, we may know the total power per discrete frequency of the white noise signal, in which case (19) would seem to be the appropriate probability density. We shall refer to the first of these two possibilities as "uniform white noise" and to the second as "Gaussian white noise."

Under some circumstances one might be willing to argue that, although we don't know the total power per discrete frequency of the noise signal, we do know an upper limit. Stated differently, not only do we know the limit x_{max} but

we know a limit for the quantity

$$P(q) = \frac{1}{n} \sum_k \langle x_k \rangle = \frac{1}{n} \int d\underset{\sim}{x} \left(\sum_k x_k \right) q(\underset{\sim}{x}) ,$$

namely

$$P(q) \leq P_{max} \quad . \tag{25}$$

As mentioned in Section II, the procedure in this case is to see if (25) is satisfied without imposing a specific constraint, and, if not, to impose the equality constraint $P = P_{max}$. The inequality constraint (25) will be satisfied by the uniform prior $p(\underset{\sim}{x})$ if

$$(x_{max}/2) \leq P_{max}. \tag{26}$$

If (26) does not hold, it follows from (19) that the appropriate density is

$$q(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/P_{max}) \exp(-x_k/P_{max}) \quad . \tag{27}$$

If x_{max} reflects knowledge about some kind of physical limit while P_{max} reflects knowledge about power limitations of the signal source, then it seems likely that (26) won't be satisfied, which means that the Gaussian white noise density (27) should be used.

IV. CROSS-ENTROPY BETWEEN INPUT AND OUTPUT OF LINEAR FILTER

Suppose a signal with probability density $p(\underset{\sim}{x})$ is passed through a linear filter with characteristic function $Y(\nu)$. Then the magnitude x_k of each power spectrum component is changed by the factor

$$S_k = \left|Y(\nu_k)\right|^2 ,$$

where $\nu_k = \omega_k/2\pi$. If $q(\underset{\sim}{x})$ is the probability density of the signal that results from passing $p(\underset{\sim}{x})$ through the filter, then the input $p(\underset{\sim}{x})$ and the output $q(\underset{\sim}{x})$ are related by

$$q(\underset{\sim}{x}) = \frac{p(x_1/S_1,\ x_2/S_2,\ \ldots,\ x_n/S_n)}{S_1 S_2 \cdots S_n} .$$

The filter has the effect of a linear coordinate transformation. The cross-entropy between the input and the output is

$$\begin{aligned} H(q,p) &= \int d\underset{\sim}{x}\ q(\underset{\sim}{x}) \log(q(\underset{\sim}{x})/p(\underset{\sim}{x})) \\ &= \int d\underset{\sim}{y}\ p(y_1,\ldots,y_n) \log\left(\frac{p(y_1,\ldots,y_n)}{p(y_1 S_1,\ldots,y_n S_n)}\right) \\ &\quad - \sum_k \log(S_k) , \end{aligned} \tag{28}$$

where $y_k = x_k/S_k$.

Eq. (28) is a general result for any input signal $p(\underset{\sim}{x})$. Now we evaluate (28) for the special case in which $p(\underset{\sim}{x})$ is uniform and for the special case in which $p(\underset{\sim}{x})$ is exponential. When the filter input $p(\underset{\sim}{x})$ is uniform, the first term in (28) is zero, and the cross-entropy between input and output is

$$H(q,p) = -\sum_k \log(S_k) \ . \tag{29}$$

Notice that, except for sign, this is the discrete form of the expression (1). When the filter input has the exponential form (21), which in terms of the spectral amplitudes a_k, b_k is Gaussian, the cross-entropy (28) becomes

$$H(q,p) = -\int d\underset{\sim}{y}\ p(\underset{\sim}{y}) \sum_k (y_k - y_k S_k)/P_k$$

$$- \sum_k \log(S_k),$$

or

$$H(q,p) = -\sum_k (1 - S_k + \log(S_k)) \ , \tag{30}$$

since $\int d\underset{\sim}{y}\ y_k p(\underset{\sim}{y}) = P_k$. Notice that this result is independent of the particular P_k values.

V. MINIMUM CROSS-ENTROPY POWER SPECTRA GIVEN AUTOCORRELATION INFORMATION

Let some unknown signal $q^+(\underset{\sim}{x})$ have a power spectrum $G(f)$ and autocorrelation function $R(t)$. Suppose we obtain information about G in the form of a set of samples of the autocorrelation function $R(t_r)$,

$$R_r = R(t_r) = \int_{-W}^{W} df\ G(f)\exp(2\pi i t_r f) \ , \tag{31}$$

$r = 1,\ldots,m$. We do not assume that the t_r are equally spaced. If the frequency spectrum is discrete, as we have assumed in (15), we can express $G(f)$ as

$$G(f) = \sum_{k=-n}^{n} G_k \delta(f - f_k) \ ,$$

where $f_k = -f_{-k}$, $G_k = G_{-k} = G(f_k)$, and $G_0 = 0$. Then (31) becomes

$$R_r = \sum_{k=-n}^{n} G_k \exp(2\pi i t_r f_k),$$

which we prefer to express in the non-complex form

$$R_r = \sum_{k=1}^{n} G_k c_{rk} , \tag{32}$$

where

$$c_{rk} = 2 \cos(2\pi t_r f_k). \tag{33}$$

Since the G_k satisfy

$$G_k = \langle x_k \rangle = \int d\underset{\sim}{x}\, x_k q^{\dagger}(\underset{\sim}{x}) , \tag{34}$$

we can rewrite (32) as

$$R_r = \int d\underset{\sim}{x}\, (\sum_k x_k c_{rk}) q^{\dagger}(\underset{\sim}{x}) . \tag{35}$$

This has the form of known expected values of the unknown density $q^{\dagger}(\underset{\sim}{x})$, and we may therefore use the principle of minimum cross-entropy to infer an estimate of $q^{\dagger}$. In terms of the general form (6), the functions g_r are $g_r = \sum_k x_k c_{rk}$. This minimum cross-entropy problem differs from the one discussed in Section III in that the Section III problem assumed knowledge of the expected spectral powers in the form (34), whereas in this problem we have only the form (35). Since typically $m < n$, knowledge of (35) provides less information than does (34).

A. Results When No Prior Power Spectrum Estimate is Given

If we have no prior information about $q^{\dagger}$, then we use a uniform prior $p(\underset{\sim}{x})$ as discussed in Section III. We then select a posterior $q(\underset{\sim}{x})$ by minimizing cross-entropy subject to the autocorrelation constraints (35) and

the normalization constraint (5). The result is

$$q(\underset{\sim}{x}) = A \exp\left(-\sum_{r=1}^{m} \beta_r \sum_{k=1}^{n} x_k c_{rk}\right) , \tag{36}$$

where the β_r are m Lagrangian multipliers corresponding to the autocorrelation constraints (35). For convenience, we define

$$u_k = \sum_{r=1}^{m} \beta_r c_{rk} , \tag{37}$$

so that (36) can be written as

$$q(\underset{\sim}{x}) = A \exp\left(-\sum_{k=1}^{n} u_k x_k\right) . \tag{38}$$

Solving for A in the usual way yields

$$q(x) = \prod_k u_k \exp(-u_k x_k) . \tag{39}$$

For our posterior estimate Q_k of the power spectrum, we use the density (39) to compute $Q_k = \langle x_k \rangle = (1/u_k)$, or

$$Q_k = \frac{1}{\sum_r \beta_r c_{rk}} , \tag{40}$$

where the multipliers β_r are determined by the constraints (32),

$$R_r = \sum_{k=1}^{n} \left[\frac{c_{rk}}{\sum_j \beta_j c_{jk}}\right] . \tag{41}$$

The minimum cross-entropy power spectrum estimate (40) is identical to the result for MESA, except that the MESA equations are usually expressed in complex form (/10, p. 9/, for example). In fact, one can derive (40) using a filtering argument of the kind usually used in deriving the MESA result. If a

white signal $p(\underset{\sim}{x})$ is passed through a linear filter with characteristic function Y, the power spectrum of the output signal is given by $Q_k = \left|Y(f_k)\right|^2$. If the output signal is to be an estimate of $q^{\dagger}$, then we know that the Q_k must satisfy (32),

$$R_r = \sum_{k=1}^{n} Q_k c_{rk} \quad . \tag{42}$$

This suggests that we design a filter with minimum cross-entropy between input output, provided that the output power spectrum satisfies (42). Since one interpretation of cross-entropy /40/ is as a measure of the information necessary to transform the prior into the posterior, one can think of this filter as the one that produces the smallest information change from the prior while still accounting for all known information. For a uniform white prior $p(\underset{\sim}{x})$, the cross-entropy between the input and output of the filter is given by (29),

$$H(q,p) = -\sum_k \log(Q_k) \tag{43}$$

Hence, we minimize (43) subject to the constraints (42). The minimization condition is

$$-(1/Q_k) + \sum_{r=1}^{m} \beta_r c_{rk} = 0 \ ,$$

and its solution for Q_k is the same as (40). Furthermore, minimizing (43) subject to (42) is just the discrete version of maximizing (1) subject to (2), which also shows the equivalence between MESA and minimum cross-entropy estimation for uniform priors. This equivalence is not surprising, since cross-entropy minimization is equivalent to entropy maximization in general for the case of uniform priors /1/.

B. Results When A Prior Power Spectrum Estimate is Given

Now we consider the case in which we obtain the autocorrelation information (35) when we already have an estimate P_k of the power spectrum G_k (34). In this case the prior density is not necessarily uniform, as it must reflect the information in the prior estimate of the power spectrum. The appropriate prior density is the exponential form (21),

$$p(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/P_k) \exp(-x_k/P_k) \ , \tag{44}$$

which itself is the minimum cross-entropy density, with respect to a uniform prior, given knowledge of the expected spectral powers P_k. If the prior estimate P_k is the estimate Q_k that was obtained by the method discussed in the previous subsection, then the appropriate prior density for obtaining a new estimate given new autocorrelation information is the posterior (39). Since $u_k = 1/Q_k$, (39) is equivalent to (44).

We therefore solve the problem of estimating G_k, given a prior estimate P_k and new autocorrelation information (35), by assuming the prior density (44) and minimizing cross-entropy subject to the constraints (35) and (5). The result is

$$\begin{aligned} q(\underset{\sim}{x}) &= p(\underset{\sim}{x}) \exp(-\lambda - \textstyle\sum_k u_k x_k) \\ &= e^{-\lambda} \prod_k \frac{1}{P_k} \exp(-(u_k + \frac{1}{P_k})x_k) \ , \end{aligned} \tag{45}$$

where the u_k are defined as in (37). Since λ's value must be such that $q(\underset{\sim}{x})$ satisfies the normalization constraint (5), (45) becomes

$$q(\underset{\sim}{x}) = (u_k + \frac{1}{P_k}) \exp(-(u_k + \frac{1}{P_k})x_k) \ . \tag{46}$$

For our posterior estimate Q_k of the power spectrum, we use the density (46)

to compute $Q_k = \langle x_k \rangle = 1/(u_k + P_k^{-1})$, or

$$Q_k = \frac{1}{(1/P_k) + \sum_r \beta_r c_{rk}} , \tag{47}$$

where as usual the multipliers β_r are determined by the requirement that the Q_k satisfy the autocorrelation constraints (42).

We can also derive the result (47) by a filtering argument that is similar to the one given in the previous subsection for the case when no previous estimate was available. Suppose a signal with power spectrum P_k is passed through a linear filter with characteristic function Y(f). The output power spectrum will be $Q_k = P_k S_k$, where $S_k = |Y(f_k)|^2$. If the output power spectrum is to be our new estimate, we know that the Q_k must satisfy (42). If the P_k is our previous estimate, this suggests that we design a filter with minimum cross-entropy between input and output, given that the input density satisfies $\langle x_k \rangle = P_k$ and that the output power spectrum Q_k satisfies (42). For input densities of the exponential form (44), the cross-entropy between input and output is given by (30). Hence, we choose the S_k so as to minimize (30) subject to the constraints (42), which we rewrite as

$$R_r = \sum_{k=1}^{n} P_k S_k c_{rk} .$$

The minimization condition is

$$1 - (1/S_k) + \sum_{r=1}^{m} \beta_r P_k c_{rk} = 0 .$$

Solving this for S_k and computing $Q_k = P_k S_k$ yields our previous result (47).

VI. COMPARISON OF RESULTS

Given information in the form of autocorrelation samples (35), the minimum cross-entropy signal probability density has the form

$$q(\underset{\sim}{x}) = \prod_{k=1}^{n} (1/Q_k) \exp(-x_k/Q_k)$$

where the Q_k are the posterior estimates of the signal power spectrum. In the case of a uniform white prior, the Q_k are given by (40),

$$Q_k^{(1)} = \frac{1}{\sum_r \beta_r^{(1)} c_{rk}} \quad . \tag{48}$$

In the case of a Gaussian white prior, the Q_k are given by (47) with $P_k = P$ for all k,

$$Q_k^{(2)} = \frac{1}{(1/P) + \sum_r \beta_r^{(2)} c_{rk}} \quad . \tag{49}$$

As discussed in Section III, P is the known value or maximum value of the expected power per unit frequency. In the case of a Gaussian non-white prior, the Q_k are given by (47),

$$Q_k^{(3)} = \frac{1}{(1/P_k) + \sum_r \beta_r^{(3)} c_{rk}} \quad , \tag{50}$$

where the P_k are prior estimates of the power spectrum. In all three cases (48)-(50), the m Lagrangian multipliers β_r are determined by the requirement that the $Q_k^{(i)}$ satisfy the autocorrelation constraints

$$R_r = \sum_{k=1}^{n} Q_k^{(i)} c_{rk} \quad , \tag{51}$$

for $r = 1,\ldots,m$ and $i = 1,2,3$.

We begin by comparing the results (48) for a uniform white prior with those (49) for a Gaussian white prior. Suppose that one of the autocorrelation samples, say R_1, is for zero lag ($t_1 = 0$). Then (33) shows that $c_{1k} = 2$ holds for all k. It follows that (49) can be written as

$$Q_k^{(2)} = \frac{1}{\sum_r \beta_r c_{rk}},$$

where $\beta_r = \beta_r^{(1)}$ for $r = 2,\ldots,m$, and $\beta_1 = \beta_1^{(1)} + (1/2P)$. Comparing with (48), we conclude that $Q_k^{(1)} = Q_k^{(2)}$ for all k. Thus, a uniform white prior and a Gaussian white prior yield the same posterior power spectrum estimates when one of the autocorrelation samples is for zero lag. This is reasonable since the zero-lag sample is just the total expected power per discrete frequency, and since the Gaussian white prior is the result of minimizing cross-entropy with respect to the uniform white prior given knowledge of the total expected power per unit frequency (See Section III).

On the other hand, the two results (48)-(49) are not equivalent in general if there is no zero-lag autocorrelation sample. To see this, suppose there is only one autocorrelation sample, $R_1 = \sum_k Q_k c_{1k}$, with $t_1 \neq 0$. If $Q_k^{(1)} = Q_k^{(2)}$ were to hold for all $k = 1,\ldots,n$, then

$$\beta_1^{(1)} c_{1k} = P^{-1} + \beta_1^{(2)} c_{1k} \tag{51}$$

would have to hold for all k. But $\beta_1^{(1)}$ and $\beta_1^{(2)}$ are constants, whereas the c_{1k} vary with k since $t_1 \neq 0$ (see (33)). It follows that (51) and, therefore, $Q_k^{(1)} = Q_k^{(2)}$ cannot hold for all k.

Now we compare the results (49) for a Gaussian white prior with the

results (50) for a Gaussian non-white prior. We consider the case of a single autocorrelation sample $R_l = \sum_k Q_k c_{lk}$ that may or may not be a zero-lag sample. If $Q_k^{(2)} = Q_k^{(3)}$ were to hold for all k = 1,...,n, then

$$P^{-1} + \beta_l^{(2)} c_{lk} = P_k^{-1} + \beta_l^{(3)} c_{lk} \tag{52}$$

would have to hold for all k. Since $\beta_l^{(2)}$ and $\beta_l^{(3)}$ are constants, and since P_k^{-1} varies with k independently of c_{lk}, (52) cannot hold for all k whether or not c_{lk} is a constant (zero-lag sample).

The results of the foregoing comparisons may be summarized as follows:

1) The results for a uniform white prior and a Gaussian white prior will be the same whenever one of the known autocorrelation samples is for zero lag. Since (48) is just the MESA result, another way of saying this is that minimum cross-entropy spectral analysis and MESA are equivalent if there is no prior estimate (other than uniform) of the power spectrum and if one of the autocorrelation samples is for zero lag.
2) If there is no zero-lag autocorrelation sample, the results for a uniform white prior and a Gaussian white prior will not in general be the same.
3) The results for a Gaussian non-white prior differ in general from those of a Gaussian white prior and those of a uniform white prior, whether or not one of the autocorrelation samples is for zero lag.

VII. TOWARDS EFFICIENT ALGORITHMS

Minimum cross-entropy spectral estimates based on autocorrelation samples are given by (48)-(50), the particular form depending on what is known about the signal prior to obtaining the autocorrelation samples. In all three cases, one must solve for the Lagrangian multipliers β_r in order to obtain actual power spectrum values Q_k. This can be done by substituting whichever of (48)-(50) is applicable into (51) and solving the m resulting equations for the β_r. Unfortunately, this approach is unlikely to be suitable for real-time signal processing although it has the advantage of considerable generality. In particular, the results (48)-(50) and (51) make no assumptions about the number and spacing of either frequencies or autocorrelation samples.

In many real-time signal processing applications, it is reasonable to assume that the the frequencies f_k are equally spaced $f_k = k\Delta f$, and that the autocorrelation samples $R(t_r)$ are equally spaced at the Nyquist rate, $t_r = 1/(2n\Delta f)$. In this case, the coefficients c_{rk} become

$$c_{rk} = 2\cos(\pi rk/n) \quad . \tag{53}$$

This form has a variety of properties that can lead to algorithmic simplifications, the fast fourier transform being the best known example. In the MESA case (48), the simplified form (53) leads to efficient algorithms that are equivalent to autoregressive and linear prediction methods /10/, /19/. It seems reasonable to expect that similar efficient techniques can also be derived from (49) and (50).

Even in cases where there the form (53) does not apply, there is one

situation in which cross-entropy spectral analysis given a prior spectral estimate and new autocorrelation samples may be particularly efficient. Suppose that the prior spectral estimates P_k are completely consistent with the new autocorrelation samples, i.e., that $R_r = \sum_k P_k c_{rk}$ holds (compare with (42) or (51)). In this case the multipliers β_r will all satisfy $\beta_r = 0$, and the posterior spectral estimates will not change from the prior estimates, $Q_k = P_k$. This follows from a general property of cross-entropy that $H(p,p) < H(q,p)$ for all $q \neq p$ (e.g., /29, p.14/), which means that the minimum cross-entropy posterior will equal the prior whenever the prior meets all constraints. Now suppose that the prior estimates P_k were obtained by cross-entropy minimization given a prior set of autocorrelation samples R'_r. The foregoing suggests that if the new samples R_r are close in value to the old samples R'_r --- i.e., the spectrum is slowly changing --- then the multipliers β_r will be close to zero. As mentioned above, the multipliers are determined in general by substituting (47) into (42) and solving the resulting equations. The result of the substitution is

$$R_r = \sum_{k=1}^{n} P_k c_{rk} \left(1 + P_k \sum_{j=1}^{m} \beta_j c_{jk}\right)^{-1} . \tag{54}$$

If the β_j are close to zero, we may expand this ignoring terms of $O(\beta_j^2)$ and higher, yielding

$$R_r = \sum_k \left(P_k c_{rk} - P_k^2 c_{rk} \sum_j \beta_j c_{jk}\right) . \tag{55}$$

However, the prior spectral estimates P_k and the prior autocorrelation samples R'_r are related by $R'_r = \sum_k P_k c_{rk}$. It follows that we

may write (55) in the form

$$\sum_{j=1}^{m} \beta_j d_{rj} = R_r - R'_r \,, \tag{56}$$

where we have defined d_{rj} as

$$d_{rj} = \sum_{k=1}^{n} P_k^2 c_{rk} c_{jk} \,. \tag{57}$$

Thus, in the case of slowly a changing autocorrelation function, (54) reduces to a set of linear equations (56), whether or not (53) holds. If (53) also holds, then the computation of (57) should be simpler, thereby making the solution of (56) even more efficient.

VIII. CONCLUSIONS

Depending on the extent to which a previous estimate is available, cross-entropy minimization provides various alternatives for estimating power spectra given autocorrelation samples. The results reduce to those of maximum entropy spectral analysis (MESA) in a manner that, in my opinion, provides a better derivation of MESA than previous approaches. When a previous power spectrum estimate P_k is available, cross-entropy minimization leads to a new estimate Q_k that differs from the MESA result. Because it exploits more information, the minimum cross-entropy estimate should be better than the MESA estimate. This possibility, and the question of efficient algorithms, will be explored in further work.

IX. ACKNOWLEDGMENT

These results were obtained during joint work with Rodney W. Johnson, whom I thank for many useful discussions.

REFERENCES

1. J. E. Shore and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," NRL Memorandum Report 3898, December, 1978.

2. J. P. Burg, "Maximum Entropy Spectral Analysis," presented at the 37th Annual Meeting Soc. of Exploration Geophysicists, Oklahoma City, Okla., 1967.

3. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, Dover, N.Y., 1959.

4. J. G. Ables, "Maximum Entropy Spectral Analysis," Astron. Astrophys. Suppl. 15, 1974, pp. 383-393.

5. E. T. Jaynes, "Information Theory and Statistical Mechanics I," Phys. Rev., 106, 1957, pp. 620-630.

6. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, Univ. of Illinois Press, Chicago, 1949.

7. A. I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.

8. E. T. Jaynes, "Prior Probabilities," IEEE Trans. on Systems Science and Cybernetics SSC-4, 1968, pp. 227-241.

9. J. P. Burg, "A New Analysis Technique for Time Series Data," presented at NATO Adv. Study Inst. on Signal Processing with Emphasis on Underwater Acoustics, 1968.

10. J. Burg, "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, 1975 (University Microfilms No. 75-25,499).

11. R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics 36, 1971, pp. 661-675.

12. T. J. Ulrych, "Maximum Entropy Power Spectrum of Truncated Sinusoids," Journal of Geophysical Research 77, 1972, pp. 1396-1400.

13. T. J. Ulrych, et. al., "Predictive Filtering and Smoothing of Short Records Using Maximum Entropy," Journal of Geophysical Research 78, 1973, pp. 4959-4964.

14. D. E. Smylie, et. al., "Analysis of Irregularities in the Earth's Rotation," Methods in Computational Physics 13, Academic Press, N.Y., 1973, pp. 391-431.

15. T. J. Ulrych and O. G. Jenson, "Cross-spectral Analysis Using Maximum Entropy," Geophysics 39, 1974, pp. 353-4.

16. T. J. Ulrych and T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Reviews of Geophysics and Space Physics, 43, No. 1, 1975, pp. 183-200.

17. S. J. Wernecke and L. R. D'Addario, "Maximum Entropy Image Reconstruction," IEEE Trans. on Computers C-26, 1977, pp. 351-364.

18. S. J. Wernecke, "Two-Dimensional Maximum Entropy Reconstruction of Radio Brightness," Radio Science 12, 1977, pp. 831-844.

19. A. van den Bos, "Alternative Interpretation of Maximum Entropy Spectral Analysis," IEEE Trans. Inf. Theory IT-17, 1971, pp. 493-4.

20. A. Wehrl, "Entropy," Rev. Mod. Phys. 50, No. 2, 1978, pp. 221-260.

21. J. S. Rowlinson, "Probability, Information, and Entropy," Nature 225, 28 March 1970, pp. 1196-1198.

22. J. M. Jauch and J. G. Baron, "Entropy, Information, and Szilard's Paradox," Helvetica Physics Acta 45, 1972, p. 220.

23. K. Friedman and A. Shimony, "Jaynes' Maximum Entropy Prescription and Probability Theory," J. Stat. Phys. 3, No. 4, 1971, pp. 381-384.

24. M. Tribus and H. Motroni, "Comments on the Paper Jaynes' Maximum Entropy Prescription and Probability Theory," J. Stat. Phys. 4, No. 2/3, 1972, pp. 227-228.

25. A. Hobson, "The Interpretation of Inductive Probabilities," J. Stat. Phys. 6, No. 2/3, 1972, pp. 189-193.

26. M. S. Bartlett, An Introduction to Stochastic Processes, Cambridge Univ. Press, Cambridge, 1966.

27. R. W. Johnson, "Computer Programs for Determining Probability Distributions by the Principles of Maximum Entropy and Minimum Cross-Entropy," to be published.

28. I. J. Good, Probability and the Weighing of Evidence, Charles Griffen, London, 1950.

29. S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.

30. I. J. Good, "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," Annals Math Stat. 34, 1963, pp. 911-934.

31. A. Hobson and B. Cheng, "A Comparison of the Shannon and Kullback Information Measures," J. Stat. Phys. 7, No. 4, 1973, pp. 301-310.

32. R. Johnson, "Axiomatic Characterization of the Directed Divergences and Their Linear Combinations," submitted to IEEE Trans. Inf. Theory.

33. E. T. Jaynes, "Information Theory and Statistical Mechanics," in Statistical Physics, V. 3 (Brandeis Lectures), (K. W. Ford, Ed.), Benjamin, Inc., New York, 1963, pp. 182-218.

34. R. S. Ingarden and A. Kossakowski, "The Poisson Probability Distribution and Information Therodynamics," Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. 9, No. 1, 1971, pp. 83-85.

35. A. Hobson, Concepts in Statistical Mechanics, Gordon and Breach, New York, 1971.

36. G. S. Arnold and J. L. Kinsey, "Information Theory for Marginal Distributions: Application to Energy Disposal in an Exothermic Reaction," J. Chem. Phys. 67, No. 8, 1977, pp. 3530-3532.

37. R. L. Kashyap, "Prior Probability and Uncertainty," IEEE Trans. Inf. Theory IT-17, 1971, pp. 641-650.

38. R. Johnson, "Comments on 'Prior Probability and Uncertainty'", IEEE Trans. Inf. Theory, to be published.

39. P. M. Lewis II., "Approximating Probability Distributions to Reduce Storage Requirements," Inf. and Control 2, 1959, pp. 214-225.

40. A. Hobson, "A New Theorem of Information Theory," J. Stat. Phys. 1, No. 3, 1969, pp. 383-391.